



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Adquisición y Transformación
Código	DTC-MBD-517
Título	Máster Universitario en Big Data por la Universidad Pontificia Comillas
Impartido en	Máster Universitario en Big Data [Primer Curso]
Créditos	3,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación

Datos del profesorado	
Profesor	
Nombre	Sofía Sánchez Urbano
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	ssurbano@icai.comillas.edu
Profesor	
Nombre	Alberto José López Espinosa
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	ajlespinosa@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura	
Aportación al perfil profesional de la titulación	
Capacidad para recoger, limpiar y estructurar grandes cantidades de datos a través de información publicada en internet.	
Recopilación de información para la creación de bases de datos.	
Prerrequisitos	
Nociones básicas de Python.	

Competencias - Objetivos	
Competencias	
Conocimientos o contenidos	
CO1	Entender los fundamentos de la analítica de datos y su aplicación en diversas áreas de la inteligencia artificial, destacando la integración en soluciones complejas y multidisciplinares para el análisis avanzado de datos masivos atendiendo a la diversidad de problemas específicos de cada área.



CO2	Comprender las técnicas de procesados de datos, las arquitecturas y herramientas más habituales y apropiadas para condiciones y requisitos de casos específicos.
Competencias	
CP1	Integrar las arquitecturas, técnicas de inteligencia artificial, análisis avanzado de datos y de visualización y de cumplimiento legal para ofrecer la solución global óptima.
CP4	Implementar las técnicas de procesado de datos y usar las herramientas más habituales y apropiadas a las condiciones y requisitos de casos específicos.
CP7	Aplicar conocimientos avanzados en Big Data y analítica de datos para desarrollar soluciones innovadoras en proyectos y en investigación, aportando y evaluando soluciones óptimas para el procesamiento y análisis de datos a gran escala.
Habilidades o destrezas	
HA1	Comunicar de manera oral y escrita con rigor técnico, claridad expositiva y coherencia argumentativa a todo tipo de interlocutores, técnicos y no técnicos.
HA2	Trabajar en equipos de carácter pluridisciplinar y/o internacional y organizar y liderar adecuadamente las dinámicas de grupo.
HA3	Desarrollar las habilidades interpersonales que requieren los entornos profesionales actuales (empatía, tolerancia, respeto, capacidad para aunar intereses contrapuestos).
HA4	Gestionar, organizar y planificar adecuadamente el trabajo y el tiempo, cumpliendo objetivos y estándares de calidad.
HA5	Mantener una formación y aprendizaje continuo y adaptación a los cambios tecnológicos y científicos.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Bloque 01. Extracción y transformación de datos. Procesos ETL. Gramática y expresiones regulares. CSV + JSON.

Bloque 02. API REST. Peticiones HTTP. Parseo de respuestas con Postman y utilizando Python. Interceptar peticiones XHR.

Bloque 03. La web como fuente de información: HTML y CSS. Proyecto de webscraping. Aspectos legales webscraping. Parseo de HTML con BeautifulSoup4 y Selenium. Interactuando con JS.

Bloque 04. Limpieza y calidad de los datos extraídos. Integridad y normalización. Problemas de codificación. Extracción de datos de PDF. OCR.

Teoría:

Clase 01. Presentación.

Clase 02. Python & Requests.



- Clase 03. Peticiones.
- Clase 04. Formatos de respuesta.
- Clase 05. Web scrapping con Bs4.
- Clase 06. Fases de proyecto.
- Clase 07. Web scrapping con Selenium.
- Clase 08. Estructura de proyecto.
- Clase 09. Logueo de errores.
- Clase 10. Limpieza de datos.
- Clase 11. Repaso.
- Clase 12. Carrera profesional.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodología Presencial: Actividades

<p>Teoría:</p> <p>Presentación de los conceptos básicos por parte del profesor y recomendaciones prácticas para llevarlos a cabo.</p> <p>Se valorará la participación de los alumnos para generar entornos de debate en clase.</p> <p>Práctica.</p> <p>Introducción a la práctica asociada a cada una de las clases teóricas y seguimiento de la misma en clase.</p> <p>Se valorará la cooperación entre los alumnos para la resolución de problemas.</p>	<p>CO1, CO2, CP1, CP4, CP7, HA1, HA2, HA3, HA4, HA5</p>
---	---

Metodología No presencial: Actividades

<p>Práctica.</p> <p>Resolución de las prácticas presentadas en clase.</p>	<p>CO1, CO2, CP1, CP4, CP7, HA1, HA2, HA3, HA4, HA5</p>
---	---

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES

<p>Clases magistrales expositivas y participativas: Exposición de contenidos fundamentales por</p>	<p>Ejercicios prácticos y resolución de problemas: Sesiones prácticas con uso de software: Actividad formativa con ordenador que, bajo la guía del profesor-tutor, fomenta el</p>
--	---



parte del profesor impulsando la reflexión y participación del alumno.	aprendizaje autónomo y/o cooperativo del alumno mediante la ejecución de programas para la consecución de los objetivos marcados	
14.00	16.00	
HORAS NO PRESENCIALES		
Estudio personal: Reflexión y análisis individual de los contenidos teóricos y prácticos de las materias y/o asignaturas del Master	Trabajos: Los alumnos tendrán que hacer trabajos breves (individuales y/o en grupo), por indicación del profesor	Proyectos: Los alumnos tendrán que hacer trabajos de tamaño medio o grande (individuales y/o en grupo), por indicación del profesor
10.00	25.00	25.00
CRÉDITOS ECTS: 3,0 (90,00 horas)		

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

El uso de IA para crear trabajos completos o partes relevantes, sin citar la fuente o la herramienta o sin estar permitido expresamente en la descripción del trabajo, será considerado plagio y regulado conforme al Reglamento General de la Universidad.

Actividades de evaluación	Criterios de evaluación	Peso
Examen tipo test.	Selección de respuestas correctas entre opciones dadas. Se busca entender conceptos teóricos vistos en clase.	30
Prácticas de scripts.	Creación de scripts y peticiones individuales. Se valorará la limpieza del código, su eficacia y el nivel de documentación.	20 %
Prácticas de proyecto.	Creación de proyectos y flujos de ejecución. Se valorará la limpieza del código, su eficacia y el nivel de documentación.	50 %

Calificaciones

Para pasar el curso la puntuación del examen debe ser mayor que 5 sobre 10.

Para pasar el curso la puntuación de la media de las prácticas debe ser mayor que 5 sobre 10.

La asistencia al 85% de las clases es obligatoria para poder hacer el examen según el Artículo 93 del Reglamento General de la Universidad de Comillas.

En esta asignatura se permite el uso de IA para planificar y testear, nunca como una solución final. La IA puede utilizarse para actividades



previas a la tarea, como la lluvia de ideas, la descripción y la investigación inicial. Este nivel (nivel 2*) se centra en el uso eficaz de la IA para la planificación, la síntesis y la generación de ideas, pero las evaluaciones deben hacer hincapié en la capacidad de desarrollar y refinar estas ideas de forma independiente. Puedes utilizar la IA para planificar, desarrollar ideas e investigar. Tu presentación final debe mostrar cómo has desarrollado y perfeccionado estas ideas.

* Perkins, Furze, Roe y MacVaugh (2024). La escala de evaluación de la IA

PLAN DE TRABAJO Y CRONOGRAMA

Actividades	Fecha de realización	Fecha de entrega
Clase 01. Presentación. Clase 02. Python & Requests. Práctica 01. Postman. Práctica 02. Aemet.	semana 01	semana 03
Clase 03. Peticiones. Clase 04. Formatos de respuesta. Práctica 03. NBA. Práctica 04. Wikipedia Continentes.	semana 02	semana 04
Clase 05. Web scrapping con Bs4. Práctica 05. Wikipedia Comillas. Día 03 de noviembre. No lectivo.	semana 03	semana 05
Clase 06. Fases de proyecto. Práctica 06. Wikipedia Universidades. Día 10 de noviembre. No lectivo.	semana 04	semana 06
Clase 07. Web scrapping con Selenium. Clase 08. Estructura de proyecto. Práctica 07. CNMC. Práctica 08. Borme I.	semana 05	semana 07
Clase 09. Logueo de errores. Clase 10. Limpieza de datos.		



Práctica 09. Borme II. Práctica 10. Borme III.	semana 06	semana 08
Clase 11. Repaso. Clase 12. Carrera profesional.	semana 07	semana 09

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Presentaciones teóricas.

Documentación oficial de librerías open source: requests, bs4, selenium & chromedriver.

Bibliografía Complementaria

Web scraping with Python, Ryan Mitchel, O'reilly. ISBN: 1491985577

Python Crash Course, Eric Matthes. ISBN: 1593279280